DOCUMENT RESUME

ED 222 867 CS 006 864

AUTHOR Freedman, Elaine S.
TITLE On Improving the Format of Examination Papers.
PUB DATE Jul 82
NOTE 23p.; Paper presented at the Annual Meeting of the United Kingdom Reading Association (19th, Newcastle upon Tyne, England, July 19-23, 1982).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Academic Achievement; Higher Education; History Instruction; Questioning Techniques; *Readability; *Reading Research; Test Anxiety; *Test Construction; *Test Format; *Testing Problems; *Test Items

ABSTRACT

An experiment investigated the effects of the instruction rubric at the beginning of an examination paper and of the format of the examination questions themselves. A history test whose questions originally differed widely in format was modified so that its rubric was as clear and concise as possible and all the short answer questions were of the same format. The design of the experiment produced four versions combining the new and old rubrics with the new and old questions so that the effects of the different changes would not be confused. Statistical analysis of data from 200 subjects seemed to indicate that the original version produced the highest candidate achievement scores; however, closer study revealed that the new rubric was in fact leading to better achievement results than the original one. Although the original rubric was clearly superior in the familiar situation, when the situation was less familiar, the new rubric began to show its worth. Readability measures also showed that the new rubric was simpler and easier to understand. Analysis of candidates' choice of questions revealed that question-style may outweigh the actual subject content in the candidate's process of choice and that questions appearing earlier in an examination are more likely to be chosen. These findings indicate that exam writers should consider the effect of the way a question is couched and the importance of readability of instructions. (JL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Paper presented at the Annual Meeting of the United Kingdom Reading Association (19th, Newcastle upon Tyne, England, July 19 - 23, 1982)

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Elaine S. Freedman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ON IMPROVING THE FORMAT OF EXAMINATION PAPERS

ELAINE S. FREEDMAN

University of Leicester
School of Education

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED222867

Problems about the unfairness introduced into examinations by question choice have led to the arrangement of examination papers becoming increasingly complicated. In turn, the instructions for choosing certain numbers of questions from different sections of the paper have tended to become more complex. The danger is that more complicated instructions are producing rubrics which are in themselves harder to understand, resulting in extra stress for the candidate, and prompting him to choose questions unwisely.

The present experiment was designed to investigate the effects of the instruction rubric at the beginning of an examination paper, and of the format of the examination questions themselves.

The findings do appear to hold implications for the constructors of examinations. Firstly, if papers are to permit choice and various different combinations of questions, then the style and difficulty of those questions ought to be comparable. Secondly, if there is to be choice, the instructions for that choice should be clear. To this end, it would seem reasonable to make use of the techniques associated with improving the Readability of material.

The overall suggestion is made that the needs of the candidate, and the possible effects of the content and structure of the examination paper upon him, be considered even more actively in the drafting of examination papers - in all subjects and at all levels.

The Current Situation

Few people would probably disagree with the viewpoint that the purpose of an examination is to give the candidate a fair opportunity to demonstrate the degree of knowledge and/or skill he has acquired in a particular area. It therefore follows that he should not be hampered in that endeavour by the examination paper itself. However, problems about the unfairness introduced into examinations by question choice have led to the arrangement of examination papers becoming complicated. In turn, the instructions for choosing certain numbers of questions from different sections of the paper have tended to become more complex, creating further difficulties.

A search of the literature reveals considerable evidence that such problems

CS036864

do occur at all levels in the British examination system: from C.S.E. and G.C.E.,* at age sixteen, right up to degree level. There appear to be five major areas of difficulty, which are all inter-connected. They concern:

The language of the examination
Question choice
Differences in intrinsic difficulty of questions
The difficulty of rubrics
Candidates' anxiety

Candidates may well not understand the language in which the examination is couched. As a result they may also have difficulty in following the instructions laid out in the examination rubric and make incorrect or unwise decisions in choosing the questions to answer. Furthermore, they may unwittingly be putting themselves at a disadvantage by choosing particular questions, some of which may well be more difficult than others in the first place.

The need for a closer study of the place of language in education has indeed been advocated for some time, and the problem was highlighted by Johnstone and Cassels in 1978,[1] reporting an experiment with 4000 secondary school pupils in England and Scotland. They experimented with school science examination questions, re-writing them to make them less complex. Pupils' scores were found to improve markedly as a result of these alterations, suggesting that the question of the appropriateness of the language of examination papers is one that should be kept closely in mind when papers are constructed. However, interestingly, it is the matter of question choice that has exercised researchers most in the last few years.

How good are candidates at choosing the questions they can answer best? An enquiry by Cowan, involving engineering students and published in 1972,[2] would suggest that they are *not* good, either at selecting the best questions to answer or at choosing their best answers once they have completed the paper.

---

* Certificate of Secondary Education and General Certificate of Education.

In addition, there are several pieces of work which demonstrate that choosing the best answers is not the most crucial problem, but rather the ways in which choice of particular questions may facilitate or handicap the student's performance. For example, Willmott and Hall's finding (1975)[3] that candidates tend to answer questions in the order they appear on the examination paper, is supported by that of Francis and Owen in 1978.[4]

The problem of question choice, in itself serious, is however likely to be compounded if individual questions themselves differ in intrinsic difficulty, and this is perhaps one of the areas that should be regarded with particular concern. We may not be able to ensure that candidates will always choose the most appropriate questions, but the construction of a good examination should make certain that the questions themselves will not produce problems of inequality.

Wilson, in Northern Ireland, found in both the 1973 and 1974 G.C.E. 'Advanced'-level Physics examinations (taken at age eighteen) that choice accounted for possible maximum differences of from 45 to 59 marks, 11 - 15% of the total available.[5]

Even at first degree level, Gowenlock, McIntosh and Mackaill (1972)[6] found similar problems with the 1968 and 1969 final degree examinations in Chemistry at Heriot-Watt University. Analyses of these results showed:

> that the questions considered were not of equal difficulty, did not test objectives equally and did not discriminate equally.[7]

Their main concern, however, was not the differential difficulty of questions, but that, depending on the choice of questions, the objectives tested by the same examination could vary from student to student. The solution they put forward was aimed at a more uniform selection of objectives, by dividing the paper into several sections, each containing questions with identical objectives. Each question would be compulsory but would have alternative elements from which the candidate would choose.[8] The same idea was also put forward by Willmott in 1972 as a way out of the dilemma caused by the

differential difficulty of questions. He also suggested that the questions be presented in sections, the choice being limited within each section.[9] This does indeed seem to be a very useful way out of a difficult dilemma, but it brings, in turn, its own problems.

Although helping to resolve the matter of question choice, a more complicated arrangement of the examination paper will necessitate a more complex rubric, which may itself produce difficulties for the candidate, as Willmott and Hall themselves point out at the beginning of their book on Question Choice. Referring to one particular paper in Geography they say:

> With papers such as the latter, it may be thought that the examination questions themselves were easy to answer compared with the disentangling of the rubric; little evidence in fact was forthcoming on this point, although it is hardly likely that the candidates' state of mind would be improved by such rubrics.[10]

Although Willmott and Hall leave the problems of the rubric at this point in order to go on to the matter of question choice, they have in fact touched upon the crucial issues in rubric design; firstly, whether candidates can understand from it what they are intended to do, and secondly, how the rubric makes them feel about answering the paper: might it inhibit performance?

Francis and Owen's interviews of chemistry candidates[4] suggested to them that there were three main categories of strategies* used in the selection of examination questions, and that use of the different strategies was related to the candidates' level of anxiety.

If, as the writer believes, it is the rubric which may ultimately hold the answer to the problem of question choice, then it is with the raising of the issue of anxiety that one comes to the heart of the matter.

Several pieces of research have demonstrated that question choice and examination performance are affected by the candidate's level of anxiety.

---

* These were 'Planners', 'Rankers' and 'Rushers'[11]

For example, in 1952, S.B. Sarason, Mandler and Craighill[12] found that by manipulating the testing situation they could alter test performance for candidates showing high or low anxiety, and following their train of thought, Smith and Rockett (1958)[13] found that it was possible to relieve anxiety by the manipulation of test instructions. Also in 1958, I.G. Sarason found that under stressful instructional conditions, 'low anxious' candidates performed better than 'high anxious' ones, yet when the stress was removed from the instructional situation, there were no differences.[14]

Given these findings, the logical extension of the argument would be that if one could make the instructions, i.e. the rubric, less stress-inducing, candidates' anxiety would tend to be relieved and performance would improve.

What makes a rubric stressful? Hambleton and Traub (1974) investigated the effect of item order on performance and on stress in a mathematics test.[15] If the test began with difficult items and went on to easier questions, performance was significantly poorer and stress was greater. Furthermore, experimental work on item-order by Munz and Jacobs, in 1971, suggests that examinees given papers beginning with the easier questions have more positive feelings after the examination (that it was easier and fairer) than candidates given questions beginning with the more difficult ones, or questions given in random order of difficulty.[16]

If starting an examination with difficult *questions* is harmful, then it is not unreasonable to assume that starting with difficult *instructions* will also be harmful. It seems clear that what is important is not whether it is questions or instructions that are involved, but rather that starting with a difficult *situation* produces higher anxiety and poorer performance.

The danger is that more complicated instructions are generating rubrics which are in themselves harder to understand. This produces two main effects.

Firstly, it increases the stress on the candidate, and secondly, it means that the candidate may choose questions unwisely. There is evidence that both effects will in turn adversely affect performance; and the problem will be compounded if the questions themselves differ in intrinsic difficulty.

The investigation described in this article involved incorporating different rubrics and different question formats into four experimental examination papers, to see whether they would affect differentially the candidates' performance in the examination. The fact that the examination chosen as the basis for the experiment was the East Midland Regional Examination Board's Certificate of Secondary Education History examination should not however be taken to mean that the findings apply only to the British situation, nor particularly to History, nor even to the school-leaving age-group alone.

A description of the experiment follows in the next section of this article, but the conclusions outlined in the final section are, it is felt, relevant to examinations as a whole, irrespective of specific subject or environment.

The Experiment*

The paper chosen as the basis for the experiment was the 1978 C.S.E. History Mode 1 Syllabus C paper: 'World Affairs - 20th Century'. The purpose was to investigate

1) the instruction rubric at the beginning of the paper
2) the format of the questions themselves

by producing alternative versions of it where the rubric and/or the questions were modified in order to facilitate the candidate's task in navigating the paper itself.** The main rubric for the whole paper was manipulated; but only the twelve questions in Part One, and not the six essay questions in Part Two, were modified in format.

The purpose of revising the rubric was to render it as clear and concise as

---

* A full account of this experiment may be found in EXPLORING THE FORMAT : the 1978 History Examination (University of Leicester School of Education, EMREB C.S.E. Research Project Report No. 4, April 1980).

** See Appendix 1 for the rubrics and a sample question.

possible, and that of altering the format of the questions to a single common one was to ensure that candidates were not penalised because of their choice of questions. On the original History paper, the formats of different questions differed widely. They ranged from multiple-choice to writing paragraphs on specific topics, from one-word answers to interpreting historical maps, from filling in slots in given passages to explaining quotations. Clearly some of these tasks were likely to be more taxing than others, irrespective of the subject matter involved. It was hoped that a common format for all the questions in Part One would remove this source of inequality, so that pupils would not be disadvantaged by their choice of question. The essay questions in Part Two of the paper would remain constant throughout.

The aim of the experiment was to compare the achievement of candidates using papers with the modified 'new' rubric with that of those using papers with the original rubric; and to compare their performance using the new-style questions with that of pupils using the original questions.

The design of the experiment would have to permit this in such a way that any differences accruing from the change in *rubric* would not be confused by the modification of the *question-style*, or vice versa. In order to achieve this end, four versions of the examination paper were prepared, which would allow the comparison of the new and the original rubrics, and of the new-style and original questions, both together and separately (where the other elements would be held constant).

The content of these papers was:

I ORIGINAL RUBRIC + ORIGINAL QUESTIONS
II NEW RUBRIC + ORIGINAL QUESTIONS
III ORIGINAL RUBRIC + NEW-STYLE QUESTIONS
IV NEW RUBRIC + NEW-STYLE QUESTIONS

and the comparisons would be made as follows:

a) In order to compare the scores of candidates taking papers containing the old rubric with the scores of those taking papers containing the new revised one, one would compare results as shown overleaf:

PAPER I (ORIGINAL RUBRIC + ORIGINAL QUESTIONS)
and
PAPER II (NEW RUBRIC + ORIGINAL QUESTIONS)

Also:

PAPER III (ORIGINAL RUBRIC + NEW-STYLE QUESTIONS)
and
PAPER IV (NEW RUBRIC + NEW-STYLE QUESTIONS)

b) To compare the achievement scores of candidates taking papers containing questions in the original style of format with those of candidates taking papers containing new-style questions, the results of the following papers would be compared:

PAPER I (ORIGINAL RUBRIC + ORIGINAL QUESTIONS)
and
PAPER III (ORIGINAL RUBRIC + NEW-STYLE QUESTIONS)

Also:

PAPER II (NEW RUBRIC + ORIGINAL QUESTIONS)
and
PAPER IV (NEW RUBRIC + NEW-STYLE QUESTIONS)

In addition to these two main areas of investigation, different combinations of contrasts of the papers would allow study of other effects of the experimental materials.

c) Thus, to investigate the possible effect of the different rubrics *combined* with the different question formats, one would compare the results from Paper I with those from Paper IV.

d) Furthermore, one would be able to see the relative influence, if any, of the different rubrics on the effect of question style, by comparing the comparison of papers I and III with the comparison of Papers II and IV (see (c) above).

e) The relative influences of the different question-styles on the effect of rubric would be visible by comparing the comparison of Papers I and II with the comparison of Papers III and IV (see (a) above).

Over 200 candidates at two schools sat the experimental papers under examination conditions, with several control measures in force to ensure the

internal validity of the experiment. For example, the four different examination papers were packed in rotation (I, II, III, IV) to be given to the schools. These papers were then to be laid out in strict order so that pupils, given a free choice of where to sit, would be assigned to their examination paper completely at random.*

Secondly, all the candidates were to be warned that there were slight variations in the presentation of the examination papers, but that the content was exactly the same. One was to be careful, however, *not* to let them realise that some papers were *experimental* or novel, so that the results would not be biased in favour of the new experimental material, through the operation of the 'Hawthorne Effect'.**

In order to avoid confusion for the students about the style of examination paper they would in fact do in the summer for the official examination, one would suggest to the schools that the students be shown the different papers at the end of the examination-experiment. At this point differences could even be discussed. In this way one would not interfere with the running of the experiment, but would also not confuse the pupils with differences in the style of this paper and the summer paper.

Thirdly, before the experimental session, standard examination procedure instructions would be given to the participating schools, together with the examination papers, to make certain that the procedures were both correct and the same for both schools.

The completed examination papers were subsequently marked twice over, by two

---

* The use of random selection meant that the experimental design would conform to Campbell and Stanley's requirements[17] for True Experimental Design No. 4, perhaps the most respected design of all, where internal validity is concerned.

** The phenomenon by which a subject's performance tends to improve if he knows that he is being studied.

independent experienced examiners, and the results were analysed using Analysis of Variance statistical techniques. As the manipulations of the rubric and of the question format were directed primarily towards Part One of the paper, and in view of the largely statistically non-significant results for Part Two and the Whole Examination, the results of the comparisons of the experimental papers will deal with Part One only. (For detailed results of comparisons (a) to (e) see Appendix 2).

The comparisons of the different papers showed Paper I (ORIGINAL RUBRIC + ORIGINAL QUESTIONS) to lead to the highest candidate achievement scores, and Paper III (ORIGINAL RUBRIC + NEW-STYLE QUESTIONS) to the lowest. Overall, Paper I was also seen to result in statistically significantly higher scores than Paper II (NEW RUBRIC + ORIGINAL QUESTIONS). The all-original Paper I also tended to produce better scores than the all new Paper IV (NEW RUBRIC + NEW-STYLE QUESTIONS).

One was thus led to the immediate conclusion that the new material was, in practice, either less successful than the original material, or at least much the same in effectiveness. However, closer study revealed an interesting phenomenon. At first sight, the original rubric had appeared superior in a straight comparison with the new one, and the traditional paper was superior to the all-new one in terms of achievement results. The picture changed, however, when one took into account the questions with which the change in rubric was linked. Studying the effect of the rubric when it was connected with the traditional questions, and its effect when connected with the new-style questions,* revealed that the new rubric was in fact leading to better achievement results than the original one. When the original questions were involved (Papers I and II) the mean score for Paper I, using the original rubric, was several marks better than that for Paper II (new rubric). However, when the new-style questions were part of the papers (Papers III and IV) it was the mean score for Paper IV, with the new rubric, which was superior to Paper III (original rubric). It would appear

---

* i.e. comparing:

I	(ORIGINAL RUBRIC + ORIGINAL QUESTIONS)	with	III	(ORIGINAL RUBRIC + NEW-STYLE QUESTIONS)
II	(NEW RUBRIC + ORIGINAL QUESTIONS)	"	IV	(NEW RUBRIC + NEW-STYLE QUESTIONS)

that when new-style questions were involved, the differences in achievement engendered by the different rubrics (in favour of the original rubric) were clearly *reversed*. Although the difference was not statistically significant, the mean achievement score of candidates taking the paper including the new rubric (Paper IV) was between 2 and 3 marks better than that for those taking the paper with the original rubric (Paper III).

The full force of this finding is clear when one considers that the average superiority of the original rubric, when combined with the original questions, was about 4.25 marks. For one of the two schools involved in the experiment, the mean score for the paper with the original rubric (Paper I) was approximately 2.5 marks better than that for the paper with the new rubric (Paper II), when original questions were involved. When, however, new questions were involved, the paper with the *new* rubric (Paper IV) produced scores 2.2 marks *better* than the old rubric paper (III), an overall change of about 4.7 marks (2.5 + 2.2). For the other school, the overall change in the difference was about 8.2 marks (5.75 + 2.45).*

Although these figures themselves cannot be taken as *conclusive*, the inference would seem to be that although the original rubric was clearly superior in the familiar situation, when the situation was less familiar (with the introduction of a new format of question) the new rubric began to show its worth. It is also interesting to note (comparison (d)) that statistically significant differences in favour of the papers with ORIGINAL QUESTIONS changed to non-significant differences when the comparison involved the NEW RUBRIC.

Tests of Readability also showed the new rubric to be less complex than the original one, and to be suitable for children with a lower reading age. The concept of Readability of a passage or text centres on the interest and motivation it generates, its legibility, and the complexity of its words and sentences in relation to the reader's reading ability. Many definitions of Readability have been devised. Perhaps Harrison's summary (1974) is one of

---

* It should perhaps be pointed out that these figures are not inconsiderable, given the overall size of the mean scores involved: see Appendix 2(e).

the most helpful:

> Broadly speaking ..... readability studies are concerned with the features of a printed text which tend to make it easier or more difficult to read and understand. These features can best be seen in terms of four groups: content, format, organization and style.[18]

In the quotation above, content refers to subject-matter, format to legibility and illustrations etc., organisation to use of paragraphs and sequencing of ideas, and style to vocabulary and the complexity of sentences.

There are a considerable number of procedures and formulae for assessing the readability of texts in practical terms,[19] so that one can begin to establish whether a given text will be suitable for a child at a given stage in his school career. The lower the 'reading age' needed to read and understand a passage, the easier that passage is deemed to be.

A very similar type of measure, known as Syntactic Density,[20] was also used to compare the new and old rubrics. The concept of Syntactic Density is linked to Readability, and is based upon the idea that the more complex the sentence structure of a passage, the harder it will be to understand. These tests again showed that the new rubric was simpler and easier to understand.

Taken together, the findings concerning the achievement results and those concerning the Readability of the rubrics suggest that it is reasonable to put forward the idea that although the new rubric may not actually have been clearly superior to the original one, it might well, as a factor in the examination situation, produce favourable results. Furthermore, the fact that the students were well-acquainted with the all-original Paper I, may well have been an important factor in its success. Over the course of time during which candidates were being prepared for the public examination (of which Paper I forms a normal part) the rubric had no doubt been 'translated' and explained to the pupils by their teacher. By contrast, the new rubric would have been completely unfamiliar to them, and had to stand or fall by its own merits alone.

Perhaps one should also bear in mind that Hambleton and Traub[21] report that Sarason[14] only found differences in performance, resulting from the different instructions given, when the conditions were stressful. Smith and Rockett[13] also found that differences in instructions had no effect on performance when the subjects were not highly anxious. The candidates who took part in the present experiment were not told that they were testing out the rubrics and question-styles of the different papers until after the examination, but they may not have been under sufficient stress for achievement differences to reflect the real differences between the rubrics.

Analysis of candidates' choice of questions also produced some interesting findings. Again, although the new question-styles did not affect performance scores materially, they did influence candidates' choice of question. When the original straightforward slot-filling technique of one question was removed in favour of the new standard format, the question's popularity declined. On the other hand, when the original complex map format of another question was replaced by the new format, that question became more popular. These findings lead one to believe that the matter of question-style is an important one, and may indeed outweigh the actual subject content in the candidate's process of choice.

This analysis also showed that, for all four papers, higher percentages of candidates tended to choose the questions appearing early in Part One rather than the ones which came later, which was also true of the questions at the beginning of Part Two of the paper. This finding may perhaps relate to Francis and Owen's categorization of examinees' strategies,[11] particularly the 'Rusher' strategy, which involves answering first the first question encountered that one is able to answer. The first three questions on the paper may have tended to be so popular for this reason.

With hindsight, one realises that an attitude questionnaire administered to the candidates immediately after the examination might have yielded

considerable information about the psychological effects the rubrics and question-styles may have had. The writer strongly suspects that such a questionnaire might well have disclosed that candidates felt somewhat less stressed by the slightly less formal instructions and the more informative rubric (as work by Shapland[22] would suggest). If the less complex new rubric did reduce stress, then that in itself would have been a valuable outcome, even if actual performance were not greatly altered in the short term. Indeed, the reduction of stress may actually aid performance in the long run.

To summarize, the results of this experiment have shown that the original paper produced better achievement scores in straight comparisons with the revised versions. However, closer study revealed use of the new rubric to be associated with superior performance when connected with the new-style questions in a more unfamiliar situation. Readability and Syntactic Density tests also showed the new rubric to be less complex than the original one, and to be suitable for children with a lower reading age. Although the new question-styles did not affect performance scores materially, they did influence candidates' choice of questions.

Suggestions for the Future

What implications do these findings hold for the constructors of examinations? Firstly, as many previous writers have pointed out, if we are going to produce papers which permit choice and various different combinations of questions, then we must ensure that the style and difficulty of the questions are comparable.

As our recent experiment demonstrated, the way in which the question was couched had an important influence on candidates' choice of questions, irrespective of the question subject-matter. We must surely be careful not to dissuade a candidate from a question, to which he knows the answer, by the

appearance of that question.

Secondly, if there is to be choice, then the instructions for that choice must be clear. The rubric should not provide a hurdle for the candidate to survive before he can begin to answer the questions which form the point of the examination paper. We must not forget Willmott and Hall's comment[10] about a particular examination being easy compared with its rubric.

Perhaps it will not be practicable, or even desirable, to submit to Readability testing every rubric and every question for every examination that is written in the future, in order to assess its suitability. However, it *would* seem to make good sense for us to apply the techniques for producing 'readable' material to the preparation of examination papers.

Clearly, it is important to think of the 'interest and motivation .... legibility .... complexity'[23] of the material, and to consider the 'content, format, organization and style'[18] of the examination papers that are prepared for candidates. It is surely essential that papers be constructed for pupils in such a way that they will be able to understand, without difficulty, what is required of them. Even if simplified rubrics and more systematically organised questions did not, in the case of this experiment, influence performance significantly, one would suggest that they are likely to be valuable in reducing the stress of examinations on candidates, and may in the long run aid performance.

It is perhaps the results of the experiment concerned with the Readability of the examination instructions which hold the most significance for those who design examination papers, as the options for study in school increase, and the consequent choice in papers becomes wider. In Britain especially, with the proposed change in emphasis from norm-referenced to criterion-referenced examinations, less able candidates are likely to have a greater opportunity to display their knowledge. It will therefore become particularly important

for them to be given examination instructions which are clear, straightforward, simple, and suitable for those with lower reading ages.

The task of the examination constructor is a very skilled one, necessitating as it does, not only the drafting of questions, but adherence to the constraints imposed by needing to cover the syllabus fairly and to produce original papers year by year. In presenting these findings one would not wish to increase the difficulty of that task.

One would however like to put forward the following overall suggestion: that the needs of the candidate, and the possible effects of the content and structure of the paper upon him, be considered even more actively in the drafting and monitoring of examination papers - in all subjects and at all levels.

© Dr. Elaine S. Freedman

University of Leicester School of Education

July 1982.

REFERENCES

1. Johnstone, A. & Cassels, J., 'What's in a Word?', *New Scientist*, 18 May 1978, pp. 432 - 434.

2. Cowan, J., 'Is freedom of choice in examinations such an advantage?', *The Technical Journal*, 10 (1), February 1972, p. 31.

3. Willmott, A.S. and Hall, C.G.W., *O Level Examined: the effect of Question Choice* (London: Macmillan Education/Schools Council Research Studies, 1975), p. 157.

4. Francis, J.C. and Owen, S.J., 'Question Choice in G.C.E. Ordinary Level Chemistry: candidates' perceptions', *Educational Studies*, 4 (2), 1978, pp. 131 - 147.

5. Wilson, J.A., 'Question Choice in A-Level Physics', *Curriculum Studies*, 8 (1), 1976, p. 74.

6. Gowenlock, B.G., McIntosh, D.M. and Mackaill, A.W., 'Choice in Chemistry Examinations - An Investigation at Honours Level', *Scottish Educational Studies*, 4 (1), 1972, pp. 21 - 27.

7. *Ibid.*, p. 25.

8. *Ibid.*, p. 26.

9. Willmott, A.S., 'G.C.E. Item Analysis - Reliability through Combinations' in *British Examinations : Techniques of Analysis*, ed. D.L. Nuttall and A.S. Willmott, (Slough: National Foundation for Educational Research, 1972), pp. 87 - 88.

10. Willmott and Hall, *op. cit.*, p. 6.

11. Francis and Owen, *op. cit.*, p. 136.

12. Sarason, S.B., Mandler, G., and Craighill, P.G., 'The Effect of Differential Instructions on Anxiety and Learning', *Journal of Abnormal and Social Psychology*, 47 (2 supplement), 1952, pp. 561 - 565.

13. Smith, W.F. and Rockett, F.C., 'Test Performance as a Function of Anxiety, Instructor and Instructions', *Journal of Educational Research*, 52 (4), 1958, pp. 138 - 141.

14. Sarason, I.G., 'Effects on Verbal Learning of Anxiety, Reassurance, and Meaningfulness of Materials', *Journal of Experimental Psychology*, 56 (6), 1958, pp. 472 - 477.

15. Hambleton, R.K. and Traub., R.E., 'The Effects of Item Order on Test Performance and Stress', *The Journal of Experimental Education*, 43 (1), 1974, pp. 40 - 46.

16. Munz, D.C. and Jacobs, P.D., 'An Evaluation of Perceived Item - Difficulty Sequencing in Academic Testing', *British Journal of Educational Psychology*, 41 (2), 1971, pp. 195 - 205.

17. Campbell, D.T. and Stanley, J.C., *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1966), pp. 13 - 22.

18. Harrison, C., *Readability and School* (University of Nottingham School of Education, Schools Council Project on The Effective Use of Reading, 1974), pp. 1 - 2.

19. See: Harrison, C., *Readability in the Classroom* (Cambridge: Cambridge University Press, 1980), pp. 51 - 83, 153 - 159.

20. See: Golub, L. and Kidder, C., 'Syntactic Density and the Computer', *Elementary English*, 51 (8), 1974, pp. 1128 - 1131.

21. Hambleton and Traub , *op. cit.*, p. 41.

22. Shapland, J.M., *Language Used in Chemistry Examination Papers*, (University of Leicester unpublished study for the Diploma in Educational Studies, 1977).

23. Johnson, R.K., 'Readability', *School Science Review*, 60 (212), 1979, p. 562.

APPENDIX 1 ORIGINAL RUBRIC

I

EAST MIDLAND REGIONAL EXAMINATIONS BOARD

Certificate of Secondary Education

Candidate's School

Candidate's Name

HISTORY

SYLLABUS C – SPECIAL PAPER I

WORLD AFFAIRS -- 20th CENTURY

2¼ HOURS

	PART ONE												PART TWO						For Examiner's Use
	Section A			Section B			Section C			Section D									
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
CANDIDATES TO TICK QUESTIONS ANSWERED																			TOTAL
FOR EXAMINER'S USE																			

Candidates are required to answer SIX questions from Part 1 of the examination paper (Questions 1-12) and THREE questions from Part 2 of the examination paper (Questions 13-18).

Credit will be given for sketchmaps, diagrams and other illustrations where they are appropriate.

The numbers of the questions answered should be ticked in the grid above.

You must clearly cancel any work which you do not wish to be marked.

Time allowed: 2¼ hours plus 10 minutes reading time (during which no writing materials may be used).

All the questions in Part 1 are worth 10 marks each and all the questions in Part 2 are worth 20 marks each.

PART ONE

You must answer SIX questions from this part.

SECTION A

For examiner use only

1. (a) From the choice given, write out the word or phrase which correctly completes the following statements:

(i) One cause of Anglo-German tension before 1914 was
the German seizure of Alsace-Lorraine.
the building of the Kiel Canal.
England's decision to end her alliance with Germany.
Germany's building of 'pocket' battleships.
the tactlessness of Kaiser William I.

.....

[P.T.O.

1V

NEW RUBRIC

EAST MIDLAND REGIONAL EXAMINATIONS BOARD

Certificate of Secondary Education

Candidate's School

Candidate's Name

HISTORY

SYLLABUS C – SPECIAL PAPER 1V

WORLD AFFAIRS – 20th CENTURY

Time allowed: 2¼ hours (plus 10 minutes reading time during which no writing materials may be used).

	PART ONE												PART TWO						For Examiner's Use
	Section A			Section B			Section C			Section D									
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
CANDIDATES TO TICK QUESTIONS ANSWERED																			Total
FOR EXAMINER'S USE																			

You are required to answer SIX questions out of the TWELVE in PART ONE of the examination paper (Questions 1 - 12), and THREE out of the SIX in PART TWO of the paper (Questions 13 - 18).

The numbers of the questions you have answered should be ticked in the grid above.

Credit will be given for sketchmaps, diagrams and other illustrations where they are appropriate.

You must cancel clearly any work you do not wish to be marked.

All the questions in Part One are worth 10 marks each and all the questions in Part Two are worth 20 marks each.

To help you decide which questions to answer in PART ONE, here is a list of the questions in that part of the paper. It shows each question number and beside it the topic covered by that question.

SECTION A	***QUESTION 1***	***The First World War.***
	QUESTION 2	***Creation of an International Organisation.***
	QUESTION 3	***First World War Peace Conferences.***
SECTION B	***QUESTION 4***	***The Spanish Civil War.***
	QUESTION 5	***The Rise of Nazi Germany.***
	QUESTION 6	***America between the World Wars.***
SECTION C	***QUESTION 7***	***The Second World War.***
	QUESTION 8	***Second World War Conferences.***
	QUESTION 9	***The United Nations.***
SECTION D	***QUESTION 10***	***China.***
	QUESTION 11	***The Cold War.***
	QUESTION 12	***The European Movement.***

20

For Examiner's Use Only

12. (a) **The Prime Minister who first took Britain into the E.E.C. was**

Harold Wilson.
Edward Heath.
Lord Home.
Anthony Eden.
Harold Macmillan.

.....**(1 mark)**

(b) **The European Statesman who tried to prevent Britain's entry into the E.E.C. was**

Adenauer.
Schmidt.
Erhardt.
Monnet.
de Gaulle.

.....**(1 mark)**

(c) **What is E.F.T.A.?**

.....

.....**(2 marks)**

(d) **Why did some British statesmen oppose Britain's entry into the E.E.C.?**

.....

.....**(2 marks)**

(e) **Write a paragraph (about four sentences) about the E.E.C.**

.....

.....

.....

.....

.....

.....

.....

.....**(4 marks)**

APPENDIX 2

RESULTS OF PAPER COMPARISONS (PART ONE TOTALS ONLY)

COMPARISON	PAPERS	FIRST EXAMINER MARKS		SECOND EXAMINER MARKS	
		MEAN SCORE	SIGNIFICANCE	MEAN SCORE	SIGNIFICANCE
a) Rubrics I & II	I *O*	26.0000	p = 0.049	26.0577	p = 0.045
	II *N*	21.7407		21.7037	
III & IV	III *O*	19.3019	Not sig.	19.7358	Not sig.
	IV *N*	21.7925		21.8491	
b) Questions I & III	I *O*	26.0000	p = 0.002	26.0577	p = 0.004
	III *N*	19.3019		19.7358	
II & IV	II *O*	21.7407	Not sig.	21.7037	Not sig.
	IV *N*	21.7925		21.8491	
c) R & Q combined I & IV	I *O*	26.0000	Not sig.	28.1923	Not sig.
	IV *N*	21.7925		22.9615	

WHERE TOTAL POSSIBLE SCORE = 60

O = ORIGINAL
N = NEW

COMPARISON d)

The effect of the different rubrics on the effect of question style

Comparison of 'I and III' with 'II and IV' (t) shows, for both schools in the experiment, significant differences in favour of paper with original questions changing to non-significant differences when comparison involved new rubric.

	SCHOOL	PAPERS	FIRST EXAMINER MARKS		SECOND EXAMINER MARKS	
			MEAN SCORE	SIGNIFICANCE	MEAN SCORE	SIGNIFICANCE
BOTH WITH ORIGINAL RUBRIC	No. 1	I *OQ*	28.0769	p = 0.036	28.1923	p = 0.038
		III *NQ*	20.6538		20.8462	
	No. 2	I *OQ*	23.9231	p = 0.020	23.9231	p = 0.037
		III *NQ*	18.0000		18.6667	
BOTH WITH NEW RUBRIC	No. 1	II *OQ*	25.4231	Not sig.	25.6923	Not sig.
		IV *NQ*	22.8846		22.9615	
	No. 2	II *OQ*	18.0357	Not sig.	18.2857	Not sig.
		IV *NQ*	20.7407		20.7778	

PART ONE TOTAL POSSIBLE SCORE = 60

COMPARISON e)

The effect of the different question-styles on the effect of rubric

Comparison of 'I and II' with 'III and IV' (a) shows significant differences in favour of paper with original rubric changing to non-significant differences when comparison involved new-style questions. In terms of actual figures, however, the differences in achievement engendered by the different rubrics (in favour of the original rubric) are clearly *reversed* for both schools' data.

	SCHOOL	PAPERS	FIRST EXAMINER MARKS		SECOND EXAMINER MARKS	
			MEAN SCORE	SIGNIFICANCE	MEAN SCORE	SIGNIFICANCE
BOTH WITH ORIGINAL QUESTIONS	No. 1	I *OR*	28.0769	Not sig.	28.1923	Not sig.
		II *NR*	25.4231		25.6923	
	No. 2	I *OR*	23.9231	p = 0.020	23.9231	p = 0.025
		II *NR*	18.0357		18.2857	
BOTH WITH NEW-STYLE QUESTIONS	No. 1	III *OR*	20.6538	Not sig.	20.8462	Not sig.
		IV *NR*	22.8846		22.9615	
	No. 2	III *OR*	18.0000	Not sig.	18.6667	Not sig.
		IV *NR*	20.7407		20.7778	

PART ONE TOTAL POSSIBLE SCORE = 60

OR = ORIGINAL RUBRIC
NR = NEW RUBRIC

N.B. The figures quoted in the text, concerning the effect of the different question-styles on the effect of the rubric at the two different schools, are based on averages of the two examiners' marks for each school.